

Docket No.: 62807-02

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

Homare KANIE, et al.

Serial No.: 10/034,991

Group Art Unit: 2171

Filed: January 03, 2002

Examiner:

For: DOCUMENT RETRIEVAL METHOD/DEVICE AND STORAGE MEDIUM
STORING DOCUMENT RETRIEVAL PROGRAM

TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Honorable Commissioner for Patents and Trademarks
Washington, D. C. 20231

Sir:


At the time the above application was filed, priority was claimed based on the
following application:

Japanese Patent Application No. 2001-002810, filed January 10, 2001

A copy of each priority application listed above is enclosed.

Respectfully submitted,

MCDERMOTT, WILL & EMERY


Keith E. George
Registration No. 34,111

600 13th Street, N.W.
Washington, DC 20005-3096
(202)756-8000 KEG:prp
Facsimile: (202)756-8087
Date: July 25, 2002

日本特許庁
JAPAN PATENT OFFICE

62807-029
Kaine et al.
10034,991

January 3, 2002

McDermott, Will & Emery

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application:

2001年 1月10日

出願番号
Application Number:

特願2001-002810

[ST.10/C]:

[JP2001-002810]

出願人
Applicant(s):

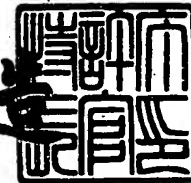
株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

2002年 1月11日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



出証番号 出証特2001-3114254

【書類名】 特許願

【整理番号】 K00012021

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 蟹江 誉

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 徳永 幹彦

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 田中 仁士

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100083552

【弁理士】

【氏名又は名称】 秋田 収喜

【電話番号】 03-3893-6221

【手数料の表示】

【予納台帳番号】 014579

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】	図面	1
【物件名】	要約書	1
【プルーフの要否】	要	

【書類名】 明細書

【発明の名称】 文書検索方法及びその実施装置並びにその処理プログラムを記録した記録媒体

【特許請求の範囲】

【請求項 1】 キーワードを用いて文書データベースから所望の文書を検索する文書検索方法において、

入力されたキーワードに関連する関連語とその関連語の有効期間を抽出するステップと、前記抽出した関連語を検索語として文書の検索を行うステップと、前記抽出した有効期間内の文書を前記検索された文書の中から選択するステップとを有することを特徴とする文書検索方法。

【請求項 2】 キーワードを用いて文書データベースから所望の文書を検索する文書検索方法において、

入力されたキーワードに関連する関連語とその関連語の有効期間を抽出するステップと、前記抽出した関連語を検索語とし、単位期間毎の検索インデックスの内、その関連語の有効期間内の検索インデックスを用いて文書の検索を行うステップとを有することを特徴とする文書検索方法。

【請求項 3】 キーワードを用いて文書データベースから所望の文書を検索する文書検索方法において、

入力されたキーワードに関連する関連語を抽出するステップと、前記抽出した関連語を検索語として文書の検索を行うステップと、前記入力されたキーワードに関連する関連語の有効期間を取得し、前記取得した有効期間内の文書を前記検索された文書の中から選択するステップとを有することを特徴とする文書検索方法。

【請求項 4】 キーワードを用いて文書データベースから所望の文書を検索する文書検索装置において、

入力されたキーワードに関連する関連語とその関連語の有効期間を抽出する時系列関連語展開処理部と、前記抽出した関連語を検索語として文書の検索を行う検索処理部と、前記抽出した有効期間内の文書を前記検索された文書の中から選択する検索結果選択処理部とを備えることを特徴とする文書検索装置。

【請求項5】 キーワードを用いて文書データベースから所望の文書を検索する文書検索装置としてコンピュータを機能させる為のプログラムを記録したコンピュータ読み取り可能な記録媒体において、

入力されたキーワードに関連する関連語とその関連語の有効期間を抽出する時系列関連語展開処理部と、前記抽出した関連語を検索語として文書の検索を行う検索処理部と、前記抽出した有効期間内の文書を前記検索された文書の中から選択する検索結果選択処理部としてコンピュータを機能させる為のプログラムを記録したことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明はキーワードを用いて文書データベースに格納されている文書から所望の文書を検索する文書検索装置に関し、特にキーワードとそのキーワードに関連する関連語の検索を行う文書検索装置に適用して有効な技術に関するものである。

【0002】

【従来の技術】

文書が大量に登録された文書データベースから所望の文書を検索する処理として全文検索がある。これはユーザの指定したキーワードが文書内に存在するものを所望の文書として検出する検索である。この検索ではユーザが任意のキーワードを指定できるが、キーワードがその関連語及び異表記によって表現された文書に関して検索に漏れが存在する問題があった。この問題を解消する為にキーワードの同義語、類義語等キーワードに関連する語も検索語として検索することにより、検索漏れを減らすという手法がある。しかしキーワードの関連語まで検索すると、検索漏れは少なくなるが、ユーザの意図とは異なる文書が検索されることも増え、ユーザの所望している文書と検索された文書との適合率が低くなるという問題があった。

【0003】

この様な問題を解決する為に、キーワードの関連語に関連の強度を設け、ユー

ザからキーワードと関連度を元に検索することで、余計な検索結果が出ない様にすることが提案されている。例えば、ユーザの意図に合った適切な関連語を得ることができ、文書検索作業をより効率的に行う文書検索装置については特開平 9 - 4 4 5 0 6 号公報に記載されている。その概要は、展開する関連語グループの関連度の範囲などの関連度条件を関連度条件入力手段により入力し、関連語同士の間連の度合いを示す間連度が関連度条件入力手段により指定された間連度条件を満たしていればその関連語グループに属する語を検索語として検索に用いるものである。

【 0 0 0 4 】

【発明が解決しようとする課題】

上記従来技術の文書検索装置では、キーワードに対して間連の強度が時間の経過と共に変化せず、一定に決まってしまうため、時間と共に類義語や間連語が変化する様なキーワードに対しての間連を行った場合、長い時間をかけて蓄積されたデータベースからは所望の文書が検索されない場合がある。また時間と共にキーワードに対して複数の間連語を登録してしまうと、検索結果に所望の文書以上の文書が含まれてしまう。

【 0 0 0 5 】

本発明の目的は上記問題を解決し、ユーザの意図に合った適切な間連語を検索して文書検索作業の効率を向上させることが可能な技術を提供することにある。

【 0 0 0 6 】

本発明の他の目的はその有効期間内の間連語の間連速度を向上させることが可能な技術を提供することにある。

【 0 0 0 7 】

本発明の他の目的は既存のシステムを大幅に変更することなく有効期間内の間連語の間連を行う構成に拡張することが可能な技術を提供することにある。

【 0 0 0 8 】

【課題を解決するための手段】

本発明は、キーワードを用いて文書データベースから所望の文書を検索する文書検索装置において、キーワードに間連する間連語の間連をその間連語の有効期

間内の文書について行うものである。

【0009】

本発明では、予めキーワードに関連する関連語とその関連語の有効期間を時系列関連語辞書に保持しておき、文書を検索しようとするユーザによってキーワードが入力されると、入力されたキーワードに関連する関連語とその関連語の有効期間を時系列関連語辞書から抽出する。そして前記抽出した関連語を検索語として文書の検索を行った後、その検索結果の文書の中から前記抽出した有効期間内の文書を選択し、前記入力されたキーワードに関連する関連語の検索結果として保持する。

【0010】

この様に本発明では、時間の経過によって類義語や関連語が変わっていくキーワードにより文書を検索する際、そのキーワード自身による検索の他に、当該キーワードから展開される同義語や類義語等の関連語の有効期間内の文書を検索して当該関連語の検索結果とするので、時間の経過に対応した適切な関連語の検索を行うことができ、ユーザの所望する文書の漏れやノイズを少なくすることができる。

【0011】

以上の様に本発明の文書検索装置によれば、キーワードに関連する関連語の検索をその関連語の有効期間内の文書について行うので、ユーザの意図に合った適切な関連語を検索して文書検索作業の効率を向上させることが可能である。

【0012】

【発明の実施の形態】

(実施形態1)

以下にキーワードに関連する関連語とその関連語の有効期間を時系列関連語辞書から抽出し、関連語を検索語とした検索結果からその関連語の有効期間内の文書を選択する実施形態1の文書検索装置について説明する。

【0013】

図1は本実施形態の文書検索装置100の概略構成を示す図である。図1に示す様に本実施形態の文書検索装置100は、CPU101と、メモリ102と、

磁気ディスク装置103と、入力装置104と、出力装置105と、CD-ROM装置106と、時系列関連語辞書130と、全文検索データベース150とを有している。

【0014】

CPU101は、文書検索装置100全体の動作を制御する装置である。メモリ102は、文書検索装置100全体の動作を制御する際にその為の各種処理プログラムやデータをロードする記憶装置である。

【0015】

磁気ディスク装置103は、前記各種処理プログラムやデータを格納しておく記憶装置である。入力装置104は、キーワードに関連する関連語を含みその関連語の有効期間内の文書を検索する為の各種入力を行う装置である。

【0016】

出力装置105は、前記文書の検索に伴う各種出力を行う装置である。CD-ROM装置106は、前記各種処理プログラムを記録したCD-ROMの内容を読み出す装置である。時系列関連語辞書130は、任意のキーワードに対する関連語とその関連語の有効期間を保持する辞書であり、関連語、有効期間、関連語を1組にしてデータを保持するものである。全文検索データベース150は、任意のキーワードまたはその関連語を含む文書とその文書を検索する為の全文検索インデクスを保持するデータベースである。

【0017】

また文書検索装置100は、キーワード入力処理部110と、時系列関連語展開処理部120と、検索処理部140と、検索結果選択処理部160と、検索結果保持処理部170とを有している。

【0018】

キーワード入力処理部110は、アプリケーション等の外部から検索の為のキーワードと検索要求を受け取る処理部である。時系列関連語展開処理部120は、キーワード入力処理部110によって入力されたキーワードに関連する関連語とその関連語の有効期間を時系列関連語辞書130から抽出する処理部である。

【0019】

検索処理部 1 4 0 は、前記抽出した関連語を検索語として全文検索データベース 1 5 0 に格納された文書の検索を行う処理部である。検索結果選択処理部 1 6 0 は、検索処理部 1 4 0 で検索された文書の作成日時と関連語の有効期間との照合を行い、前記抽出した有効期間内の文書を前記検索された文書の中から選択する処理部である。検索結果保持処理部 1 7 0 は、検索結果選択処理部 1 6 0 での選択によって得られた文書を検索結果として保持する処理部である。

【 0 0 2 0 】

文書検索装置 1 0 0 をキーワード入力処理部 1 1 0、時系列関連語展開処理部 1 2 0、検索処理部 1 4 0、検索結果選択処理部 1 6 0 及び検索結果保持処理部 1 7 0 として機能させる為のプログラムは、CD-ROM等の記録媒体に記録され磁気ディスク等に格納された後、メモリにロードされて実行されるものとする。なお前記プログラムを記録する記録媒体はCD-ROM以外の他の記録媒体でも良い。

【 0 0 2 1 】

本実施形態では、キーワードに関連する関連語を検索語とする検索について説明するが、キーワードを検索語とする検索も別途行われているものとし、他の実施形態についても同様であるものとする。

【 0 0 2 2 】

図 2 は本実施形態の検索処理の処理手順を示すフローチャートである。図 2 に示すフローチャートを元に図 1 の様に構成された本実施形態の処理について説明する。

【 0 0 2 3 】

まずステップ 2 0 1 で文書検索装置 1 0 0 のキーワード入力処理部 1 1 0 は、アプリケーション等の外部から検索の為のキーワードと検索要求を入力する。ステップ 2 0 2 で時系列関連語展開処理部 1 2 0 は、時系列関連語辞書 1 3 0 を参照して、キーワード入力処理部 1 1 0 により入力されたキーワードに一致する関連元語を探し、キーワードに一致する関連元語に対応する関連語と有効期間を抽出して、前記入力されたキーワードの有効期間情報付きの関連語一覧としてメモリに展開する。

【0024】

次にステップ203で検索処理部140は、ステップ202で展開された関連語を含む文書を全文検索データベース150から検索し、その関連語を含む文書の作成日と検索の対象となった関連語を一覧としてメモリに展開する。

【0025】

ステップ204で検索結果選択処理部160は、検索でヒットした文書数をループ回数に設定してステップ205へ進む。ステップ205では、ステップ203で検索された文書の作成日がステップ202で抽出した関連語の有効期間内であるかどうかを調べ、その文書の作成日が関連語の有効期間内であればステップ206に進む。ステップ206で検索結果保持処理部170は、その文書を一意に識別する為の文書識別子を一覧に追加して検索結果としてメモリ保持する。またその文書の作成日が関連語の有効期間内でなければステップ205に戻り、次の文書に対して同様の処理を行う。

【0026】

図3は本実施形態の検索処理の具体例を示す図である。次に、図3に示す通り具体例を用いて実際の処理内容を説明する。例えば「首相」という語をキーワードとして検索する場合を考える。

【0027】

まず、キーワード入力処理部110が「首相」というキーワード301を入力したとする。時系列関連語展開処理部120は、時系列関連語辞書130を用いて関連語と有効期間を抽出し、一覧302としてメモリに展開する。時系列関連語辞書130は、「首相」というキーワードに対して、関連語として「歴代の首相の名前」、有効期間として「歴任した期間」を保持している。また、この他に時系列関連語辞書130は、「大統領」というキーワードに対して、関連語として「歴代のアメリカ大統領の名前」、その有効期間として「歴任した期間」を保持している。ここでは「首相」というキーワードを「歴代の首相の名前」と「歴任した期間」の一覧302として展開している。

【0028】

検索処理部140は、全文検索データベース150を用いて一覧302にある

関連語を含む文書を検索する。この際に作成日時及び対象となった関連語を一覧でメモリに展開する。ここでは全文検索データベース150を検索した結果、1997.10.29に作成され対象の関連語が「橋本龍太郎」である文書0010の他、文書0001、文書0013、文書0102、文書0025、文書0123及び文書0254が一覧303として展開されている。

【0029】

検索結果選択処理部160は、一覧303に展開されているそれぞれの文書に関して文書作成日時が一覧302で取得している関連語の有効期間に一致するかどうかを判定し、一致するものを検索結果304に追加し、そうでない場合には検索結果304に含めない処理を行う。ここでは文書0010の作成日時「1997.10.29」は関連語「橋本龍太郎」の有効期間「1996.01.11-1998.07.30」に含まれているので検索結果304に追加するが、文書0013の作成日時「1997.03.03」は関連語「小淵恵三」の有効期間「1998.07.30以降」に含まれていないので検索結果304としていない。こうして得られた検索結果304を検索結果保持処理部170で保持する。

【0030】

従来の方法では、時間的に意味が変化するキーワードに対しても一定の関連語に展開して検索を行う為、ユーザが意図するものとは異なるものまで検索結果に含まれ、ユーザが所望の文書かどうかの判定作業に多大な時間を要していたが、本実施形態によれば、キーワードの時間の経過による意味の違いを意識し、展開された関連語の有効期間内の文書を検索するので、関連語の検索の際にユーザが意図しない文書が検索されることが少なくなり、検索作業の効率を向上させることが可能となる。

【0031】

以上説明した様に本実施形態の文書検索装置によれば、キーワードに関連する関連語の検索をその関連語の有効期間内の文書について行うので、ユーザの意図に合った適切な関連語を検索して文書検索作業の効率を向上させることが可能である。

(実施形態2)

以下にキーワードに関連する関連語の検索をその有効期間内の検索インデクスを用いて行う実施形態2の文書検索装置について説明する。

【0032】

図4は本実施形態の文書検索装置100の概略構成を示す図である。図4に示す様に本実施形態の文書検索装置100は、時系列関連語辞書230と、時系列全文検索データベース250とを有している。

【0033】

時系列関連語辞書230は、任意のキーワードに対する関連語とその関連語の有効期間を保持する辞書であり、関連語、有効期間、関連元語を1組にしてデータを保持するものである。時系列全文検索データベース250は、任意のキーワードまたはその関連語を含む文書と、その文書を検索する為の単位期間毎の全文検索インデクスを保持するデータベースであり、単位期間とその期間内に作成された文書への全文検索インデクスを組にしてデータを保持するものである。

【0034】

また文書検索装置100は、キーワード入力処理部210と、時系列関連語展開処理部220と、時系列検索処理部240と、検索結果保持処理部260とを有している。

【0035】

キーワード入力処理部210は、アプリケーション等の外部から検索の為のキーワードと検索要求を受け取る処理部である。時系列関連語展開処理部220は、キーワード入力処理部210によって入力されたキーワードに関連する関連語とその関連語の有効期間を時系列関連語辞書230から抽出する処理部である。

【0036】

時系列検索処理部240は、前記抽出した関連語を検索語とし、時系列全文検索データベース250の単位期間毎の検索インデクスの内、その関連語の有効期間内の検索インデクスを用いて文書の検索を行う処理部である。検索結果保持処理部260は、時系列検索処理部240での検索によって得られた文書を検索結果として保持する処理部である。

【0037】

文書検索装置100をキーワード入力処理部210、時系列関連語展開処理部220、時系列検索処理部240及び検索結果保持処理部260として機能させる為のプログラムは、CD-ROM等の記録媒体に記録され磁気ディスク等に格納された後、メモリにロードされて実行されるものとする。なお前記プログラムを記録する記録媒体はCD-ROM以外の他の記録媒体でも良い。

【0038】

図5は本実施形態の検索処理の処理手順を示すフローチャートである。図5に示すフローチャートを元に図4の様に構成された本実施形態の処理について説明する。

【0039】

まずステップ501で文書検索装置100のキーワード入力処理部210は、アプリケーション等の外部から検索の為のキーワードと検索要求を入力する。ステップ502で時系列関連語展開処理部220は、時系列関連語辞書230を参照して、キーワード入力処理部210により入力されたキーワードに一致する関連元語を探し、キーワードに一致する関連元語に対応する関連語と有効期間を抽出して、前記入力されたキーワードの有効期間情報付きの関連語一覧としてメモリに展開する。

【0040】

ステップ503で時系列検索処理部240は、ステップ502で展開された関連語数をループ回数に設定してステップ504へ進む。ステップ504では、時系列全文検索データベース250に存在する全文検索インデクス数をループ回数に設定してステップ505へ進む。

【0041】

ステップ505では、全文検索インデクスの単位期間と関連語の有効期間とを比較し、それらが重なる場合にはステップ506へ進む。ステップ506では、その全文検索インデクスを用いて当該関連語の検索を行う。ステップ507では、ステップ506での検索の結果、文書が検索されたかどうかを調べ、文書が検索された場合にはステップ508へ進む。

【0042】

ステップ508では、検索された文書数をループ回数に設定してステップ509に進む。ステップ509では、前記検索された文書の作成日時が関連語の有効期間内であるかどうかを調べ、文書の作成日時が関連語の有効期間内である場合にはステップ510へ進む。ステップ510で検索結果保持処理部260は、その文書を一意に識別する為の文書識別子を一覧に追加して検索結果としてメモリに保持する。

【0043】

ステップ509で文書の作成日時が関連語の有効期間内であるかどうかを調べた結果、文書の作成日時が関連語の有効期間内でなければ次の文書の作成日時が関連語の有効期間内であるかどうかを調べる。またステップ505で全文検索インデクスの単位期間と関連語の有効期間とを比較した結果、それらが重ならない場合には次の全文検索インデクスの単位期間について比較を行う。また全ての全文検索インデクスの単位期間についてその関連語の有効期間との比較が終了した場合には、次の関連語の有効期間について全文検索インデクスの単位期間との比較を行う。

【0044】

図6は本実施形態の検索処理の具体例を示す図である。次に、図6に示す通り具体例を用いて実際の処理内容を説明する。例えば「首相」という語をキーワードとして検索する場合を考える。

【0045】

まず、キーワード入力処理部210から「首相」というキーワード601を入力したとする。時系列関連語展開処理部220は、時系列関連語辞書230を用いて関連語と有効期間を抽出し、一覧602としてメモリに展開する。時系列関連語辞書230には、「首相」というキーワードに対して、関連語として「歴代の首相の名前」、有効期間として「歴任した期間」を保持している。また、この他に時系列関連語辞書230は、「大統領」というキーワードに対して、関連語として「歴代のアメリカ大統領の名前」、有効期間として「歴任した期間」を保持している。ここでは「首相」というキーワードを「歴代の首相の名前」と「歴任した期間」の一覧602として展開している。

【 0 0 4 6 】

時系列検索処理部 2 4 0 は、一覧 6 0 2 を元に時系列全文検索データベース 2 5 0 を用いて文書の検索を行う。例えば関連語「小淵恵三」の有効期間は「1998.07.30以降」であるから、時系列全文検索データベース 2 5 0 の期間「1998.07.30-1998.12.31」と「1999.01.01-1999.12.31」と「2000.01.01以降」の全文検索インデックスを検索する。後者の全文検索インデックス中に「小淵恵三」を含む文書 0102 が存在し、且つ文書 0102 の作成日時は「2000.03.05」であり、関連語「小淵恵三」の有効期間「1998.07.30以降」に当てはまることから、文書 0102 は所望の文書と判定し、検索結果 6 0 3 として追加する。また時系列全文検索データベース 2 5 0 の期間「1997.01.01-1997.12.31」の全文検索インデックス中にキーワード「小淵恵三」を含む文書 0013 と文書 0009 が存在するが、これは関連語「小淵恵三」の有効期間「1998.07.30以降」に当てはまらないので検索結果 6 0 3 に含まれることはない。

【 0 0 4 7 】

一覧 6 0 2 に展開されたそれぞれの関連語に対して同様の処理を行って検索結果 6 0 3 とし、検索結果保持処理部 2 6 0 で保持する。

【 0 0 4 8 】

本実施形態によれば、時系列全文検索データベース 2 5 0 の全文検索インデックスが単位期間で分かれていることからデータベース中の文書全体に検索をかけずに済み、また全文検索インデックスから検索される文書の量も全体から検索される文書の量に比べて限られるので、文書の作成日時と関連語の有効期間チェックをする回数が少なくなることから効率的な検索が行えると言える。

【 0 0 4 9 】

以上説明した様に本実施形態の文書検索装置によれば、キーワードに関連する関連語の検索をその有効期間内の検索インデックスを用いて行うので、その有効期間内の関連語の検索速度を向上させることが可能である。

(実施形態 3)

以下に関連語の有効期間を関連語有効期間データベースから取得し、キーワードに関連する関連語の検索結果からその関連語の有効期間内の文書を選択する実

施形態 3 の文書検索装置について説明する。

【0050】

図 7 は本実施形態の文書検索装置 100 の概略構成を示す図である。図 7 に示す様に本実施形態の文書検索装置 100 は、関連語辞書 330 と、全文検索データベース 350 と、関連語有効期間データベース 370 とを有している。

【0051】

関連語辞書 330 は、任意のキーワードを関連語に展開する為の関連語の集合を管理する辞書である。全文検索データベース 350 は、任意のキーワードまたはその関連語を含む文書とその文書を検索する為の全文検索インデックスを保持するデータベースである。

【0052】

関連語有効期間データベース 370 は、任意のキーワードに対する関連語の有効期間を取得する為にキーワード、関連語と有効期間の関係を管理しているデータベースであり、関連語、有効期間、関連元語を 1 組にしてデータを保持するものである。

【0053】

また文書検索装置 100 は、キーワード入力処理部 310 と、関連語展開処理部 320 と、検索処理部 340 と、検索結果選択処理部 360 と、検索結果保持処理部 380 とを有している。

【0054】

キーワード入力処理部 310 は、アプリケーション等の外部から検索の為にキーワードと検索要求を受け取る処理部である。関連語展開処理部 320 は、キーワード入力処理部 310 によって入力されたキーワードに関連する関連語を関連語辞書 330 から抽出する処理部である。

【0055】

検索処理部 340 は、前記抽出した関連語を検索語として全文検索データベース 350 に格納された文書の検索を行う処理部である。検索結果選択処理部 360 は、関連語展開処理部 320 で抽出した関連語の有効期間を関連語有効期間データベース 370 から取得し、検索処理部 340 で検索された文書の作成日時と

関連語の有効期間との照合を行い、前記取得した有効期間内の文書を前記検索された文書の中から選択する処理部である。検索結果保持処理部380は、検索結果選択処理部360での選択によって得られた文書を検索結果として保持する処理部である。

【0056】

文書検索装置100をキーワード入力処理部310、関連語展開処理部320、検索処理部340、検索結果選択処理部360及び検索結果保持処理部380として機能させる為のプログラムは、CD-ROM等の記録媒体に記録され磁気ディスク等に格納された後、メモリにロードされて実行されるものとする。なお前記プログラムを記録する記録媒体はCD-ROM以外の他の記録媒体でも良い。

【0057】

図8は本実施形態の検索処理の処理手順を示すフローチャートである。図8に示すフローチャートを元に、図7の様に構成された本実施形態の動作について説明する。

【0058】

まずステップ801で文書検索装置100のキーワード入力処理部310は、アプリケーション等の外部から検索の為のキーワードと検索要求を入力する。ステップ802で関連語展開処理部320は、関連語辞書330を参照して、キーワード入力処理部310により入力されたキーワードに関連する関連語を抽出して、前記入力されたキーワードの関連語一覧としてメモリに展開する。

【0059】

ステップ803で検索処理部340は、ステップ802で展開された関連語を含む文書を全文検索データベース350から検索し、ヒットした対象の関連語、文書の作成日時を取得する。

【0060】

ステップ804で検索結果選択処理部360は、ステップ803の検索でヒットした文書数をループ回数に設定してステップ805に進む。ステップ805では、検索を行った関連語の有効期間を関連語有効期間データベース370から取

得する。

【0061】

ステップ806では、前記取得した関連語の有効期間と文書の作成日時とを比較し、文書の作成日時が関連語の有効期間内であればステップ807へ進み、そうでなければ次の文書の作成日時が関連語の有効期間内であるかどうかを調べる。ステップ807で検索結果保持処理部380は、その文書を一意に識別する為の文書識別子を一覧に追加して検索結果としてメモリに保持する。

【0062】

図9は本実施形態の検索処理の具体例を示す図である。次に図9に示す通り具体例を用いて処理内容を説明する。例えば「首相」という語をキーワードとして検索する場合を考える。

【0063】

まず、キーワード入力処理部310から「首相」というキーワード901を入力したとする。関連語展開処理部320は、関連語辞書330を用いて「首相」というキーワードを含む関連語グループの関連語の一覧902をメモリに展開する。ここでは「首相」というキーワードに対して「歴代の首相の名前」に展開されていることを示している。検索処理部340は一覧902を元に全文検索データベース350を用いて文書の検索を行い、ヒットした文書のID、対象の関連語、作成日時を一覧903としてメモリに展開する。

【0064】

検索結果選択処理部360は、一覧903のそれぞれの文書に対して関連語有効期間データベース370から関連語の有効期間を取得し、文書の作成日時との比較を行う。例えば文書0010については、関連語有効期間データベース370から取得した関連語「橋本龍太郎」の有効期間が「1996.01.11-1998.07.30」であり、文書の作成日時「1997.10.29」が有効期間内であるので検索結果904に追加される。また、文書0013については、関連語有効期間データベース370から取得した関連語「小渕恵三」の有効期間が「1998.07.30以降」であり、文書の作成日時「1997.03.03」が有効期間内ではないので検索結果904にならない。一覧903で展開されたそれぞれの文書に対して同様の処理を行って検索結果90

4とし、検索結果保持処理部380で保持する。

【0065】

本実施形態の文書検索装置100では、前半の検索処理部340までは既にある構成を用いることが可能であり、その構成に検索結果選択処理部360及び関連語有効期間データベース370を加えることで実装することができることから、既存の構成に対して機能拡張しやすい形態であると言える。

【0066】

以上説明した様に本実施形態の文書検索装置によれば、関連語の有効期間を関連語有効期間データベースから取得し、キーワードに関連する関連語の検索結果からその関連語の有効期間内の文書を選択するので、既存のシステムを大幅に変更することなく有効期間内の関連語の検索を行う構成に拡張することが可能である。

【0067】

【発明の効果】

本発明によればキーワードに関連する関連語の検索をその関連語の有効期間内の文書について行うので、ユーザの意図に合った適切な関連語を検索して文書検索作業の効率を向上させることが可能である。

【図面の簡単な説明】

【図1】

実施形態1の文書検索装置の概略構成を示す図である。

【図2】

実施形態1の検索処理の処理手順を示すフローチャートである。

【図3】

実施形態1の検索処理の具体例を示す図である。

【図4】

実施形態2の文書検索装置の概略構成を示す図である。

【図5】

実施形態2の検索処理の処理手順を示すフローチャートである。

【図6】

実施形態 2 の検索処理の具体例を示す図である。

【図 7】

実施形態 3 の文書検索装置の概略構成を示す図である。

【図 8】

実施形態 3 の検索処理の処理手順を示すフローチャートである。

【図 9】

実施形態 3 の検索処理の具体例を示す図である。

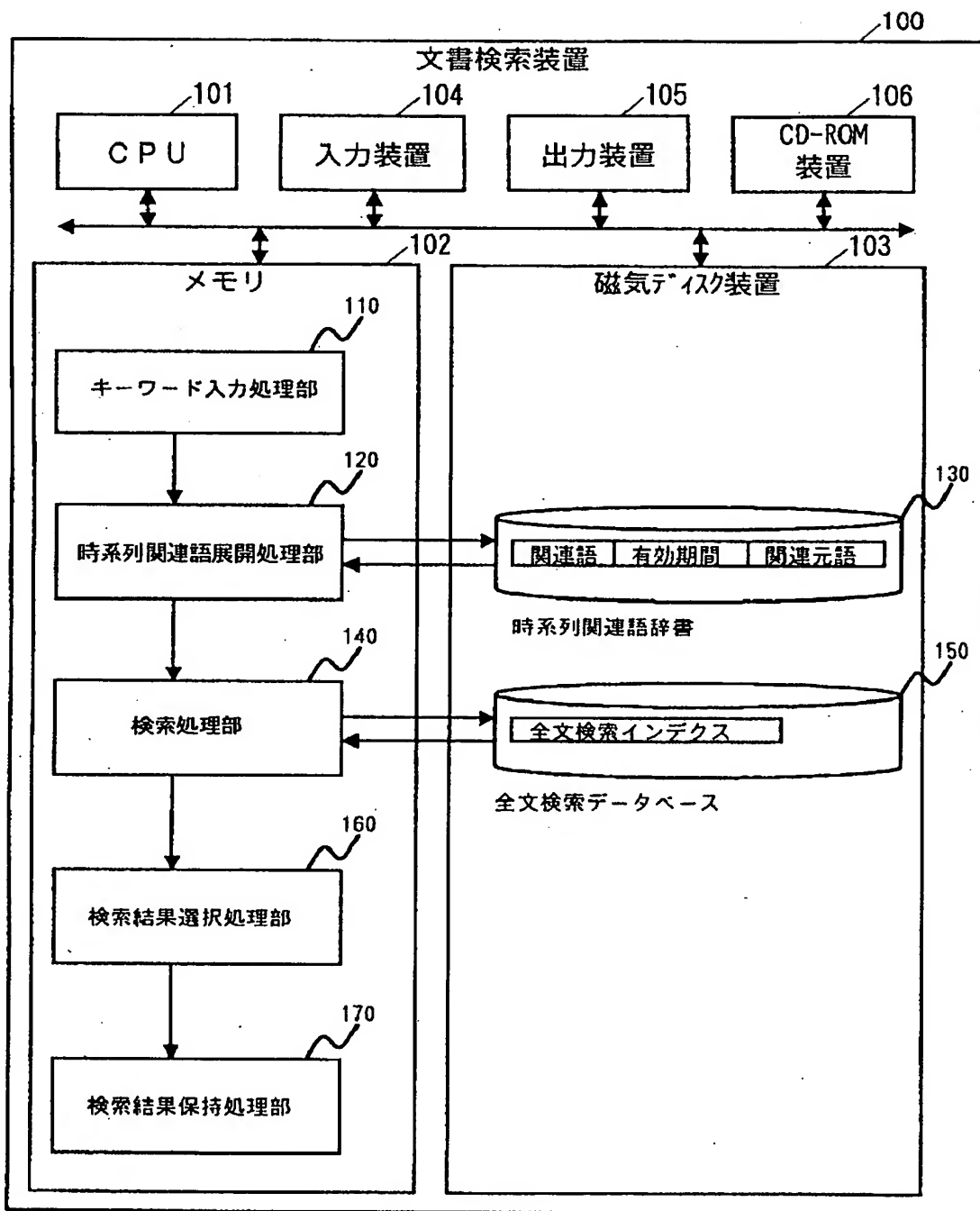
【符号の説明】

100…文書検索装置、101…CPU、102…メモリ、103…磁気ディスク装置、104…入力装置、105…出力装置、106…CD-ROM装置、130…時系列関連語辞書、150…全文検索データベース、110…キーワード入力処理部、120…時系列関連語展開処理部、140…検索処理部、160…検索結果選択処理部、170…検索結果保持処理部、301…キーワード、302～303…一覧、304…検索結果、230…時系列関連語辞書、250…時系列全文検索データベース、210…キーワード入力処理部、220…時系列関連語展開処理部、240…時系列検索処理部、260…検索結果保持処理部、601…キーワード、602…一覧、603…検索結果、330…関連語辞書、350…全文検索データベース、370…関連語有効期間データベース、310…キーワード入力処理部、320…関連語展開処理部、340…検索処理部、360…検索結果選択処理部、380…検索結果保持処理部、901…キーワード、902～903…一覧、904…検索結果。

【書類名】 図面

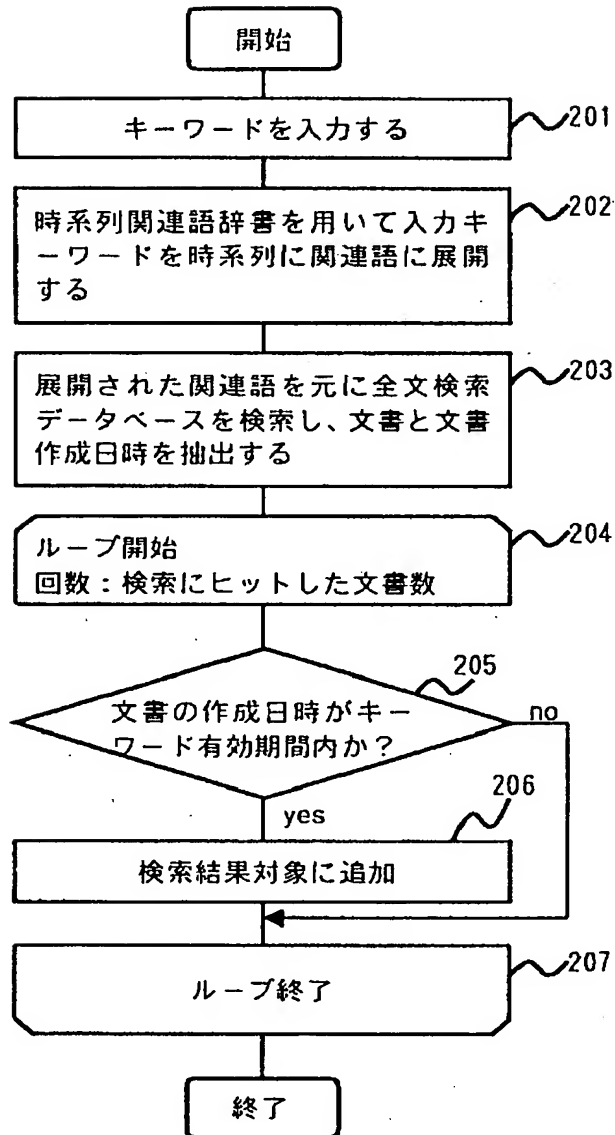
【図 1】

図 1



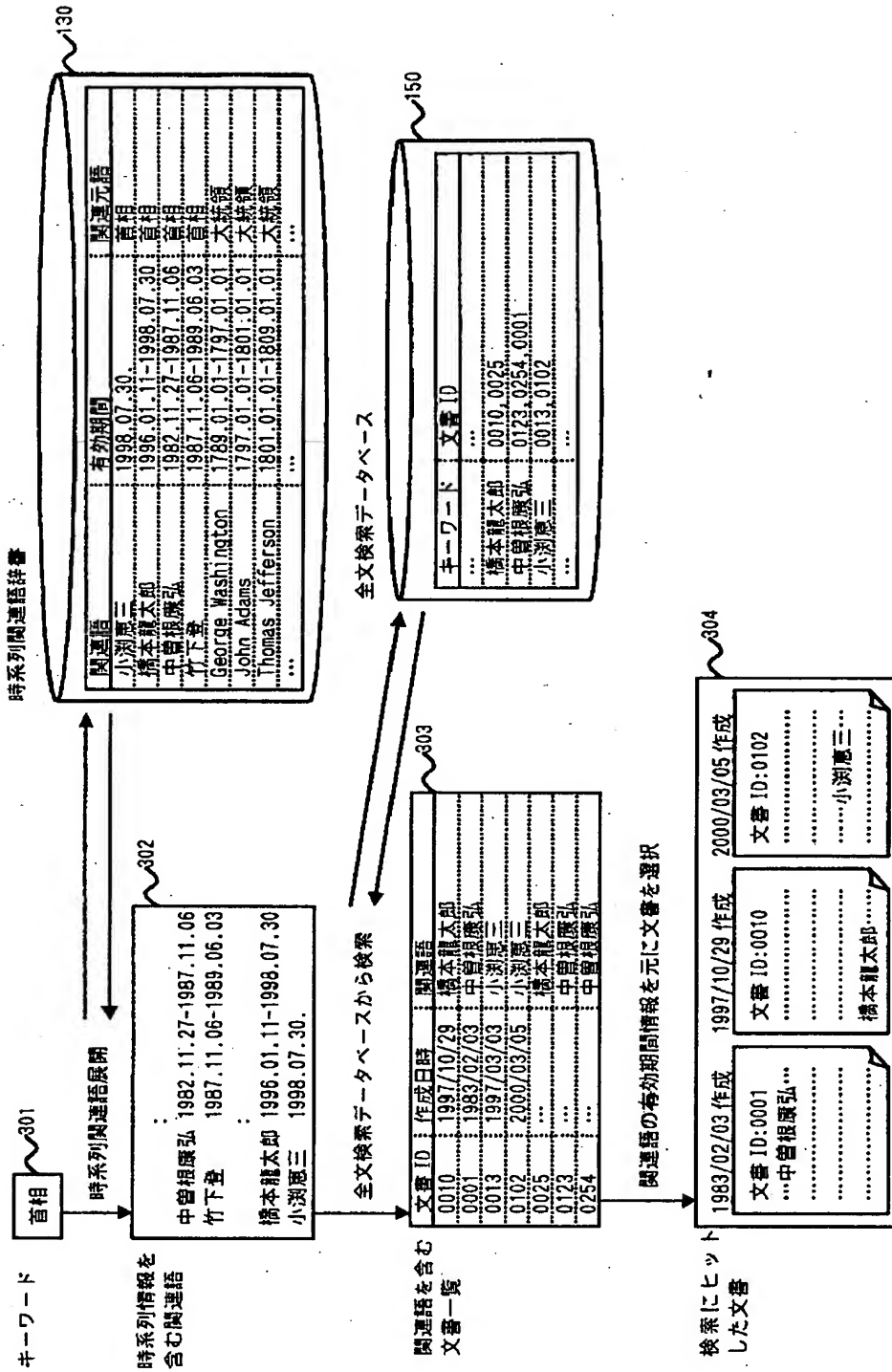
【図 2】

図 2



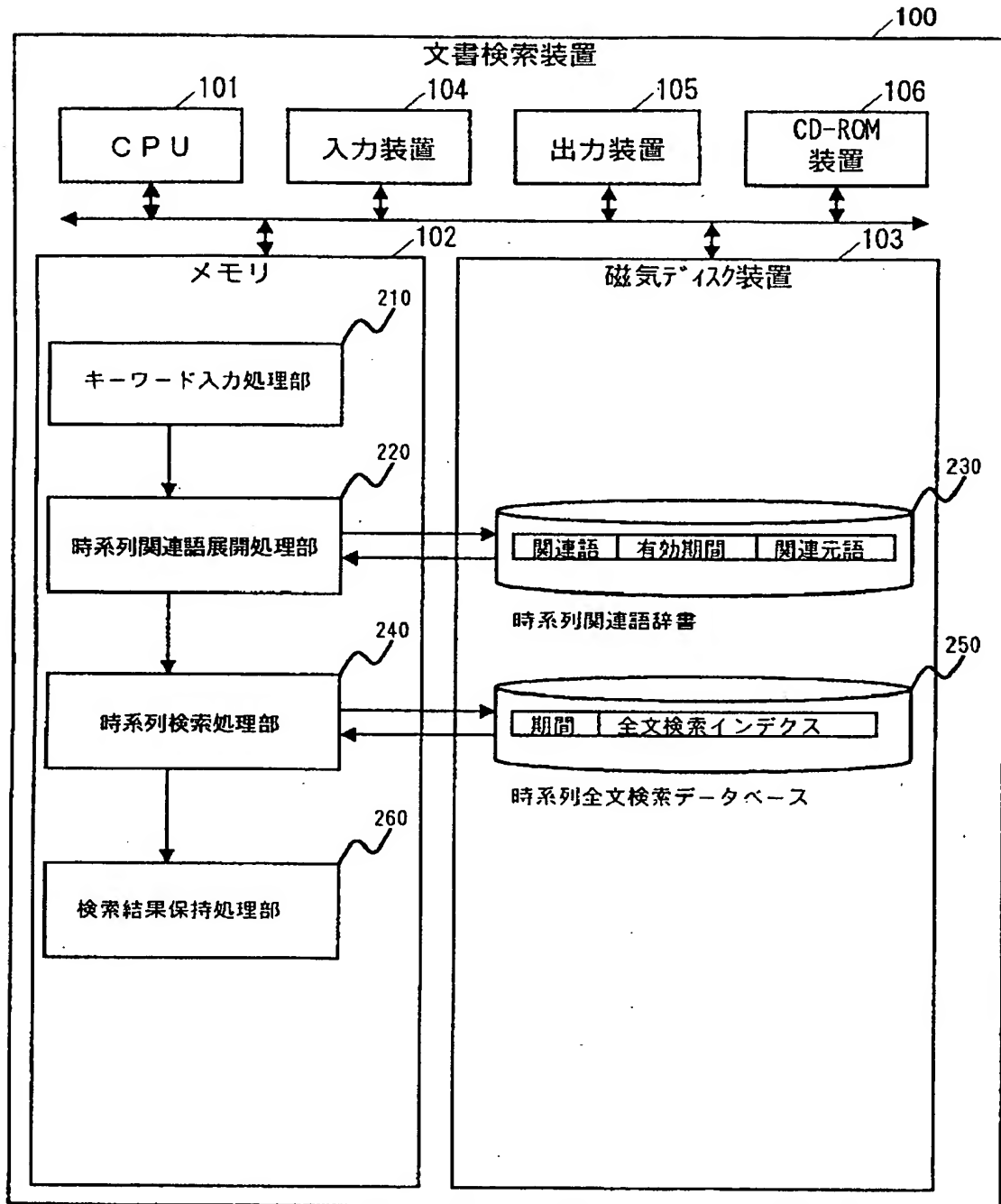
【図 3】

図 3



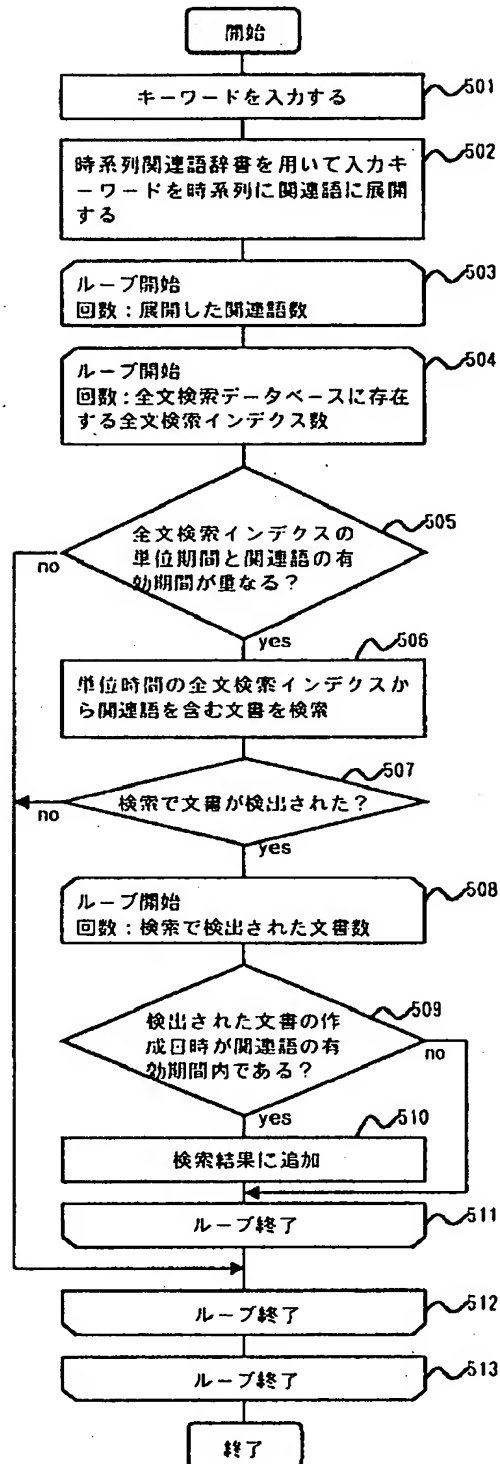
【図 4】

図 4



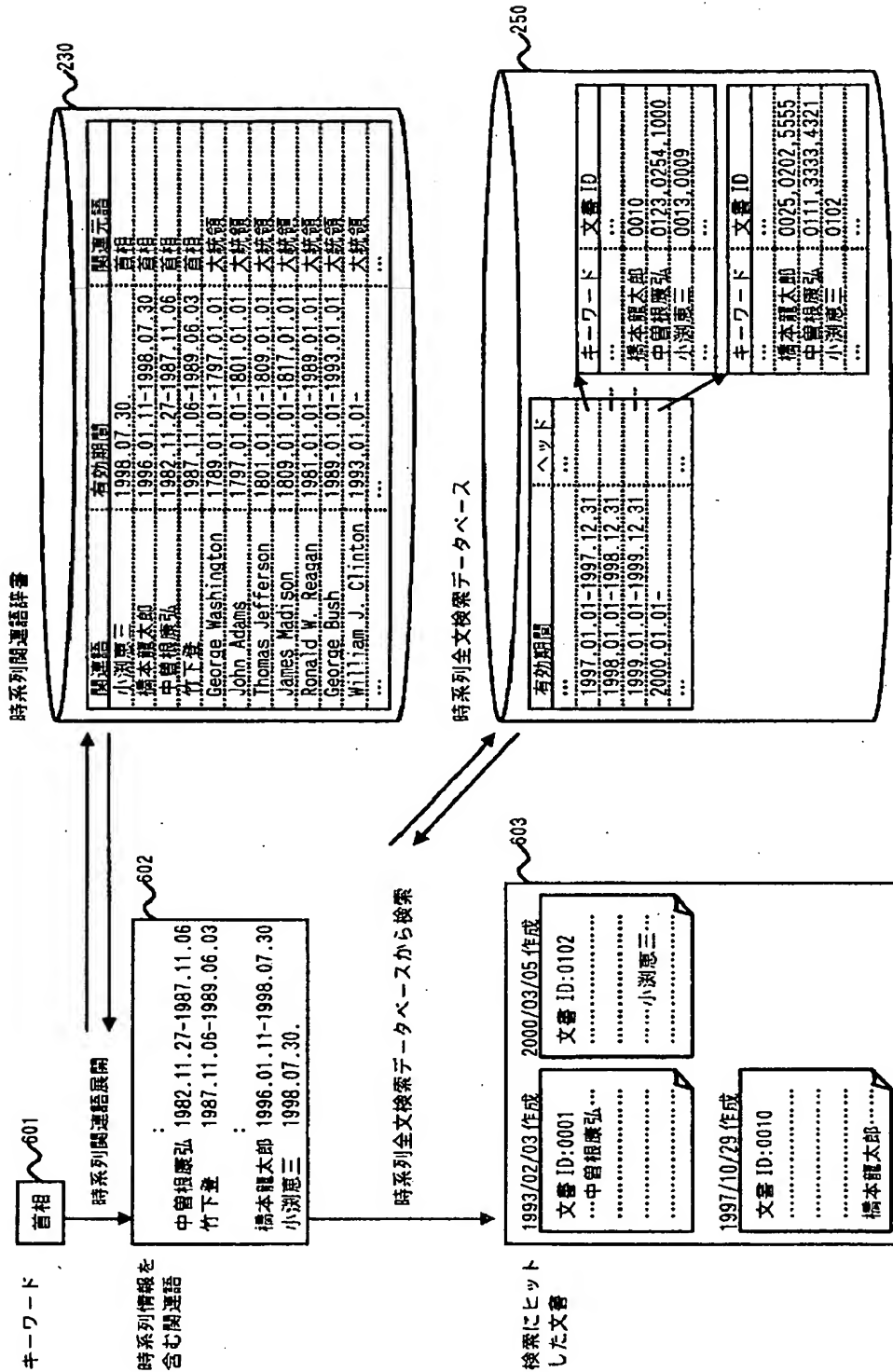
【図 5】

図 5



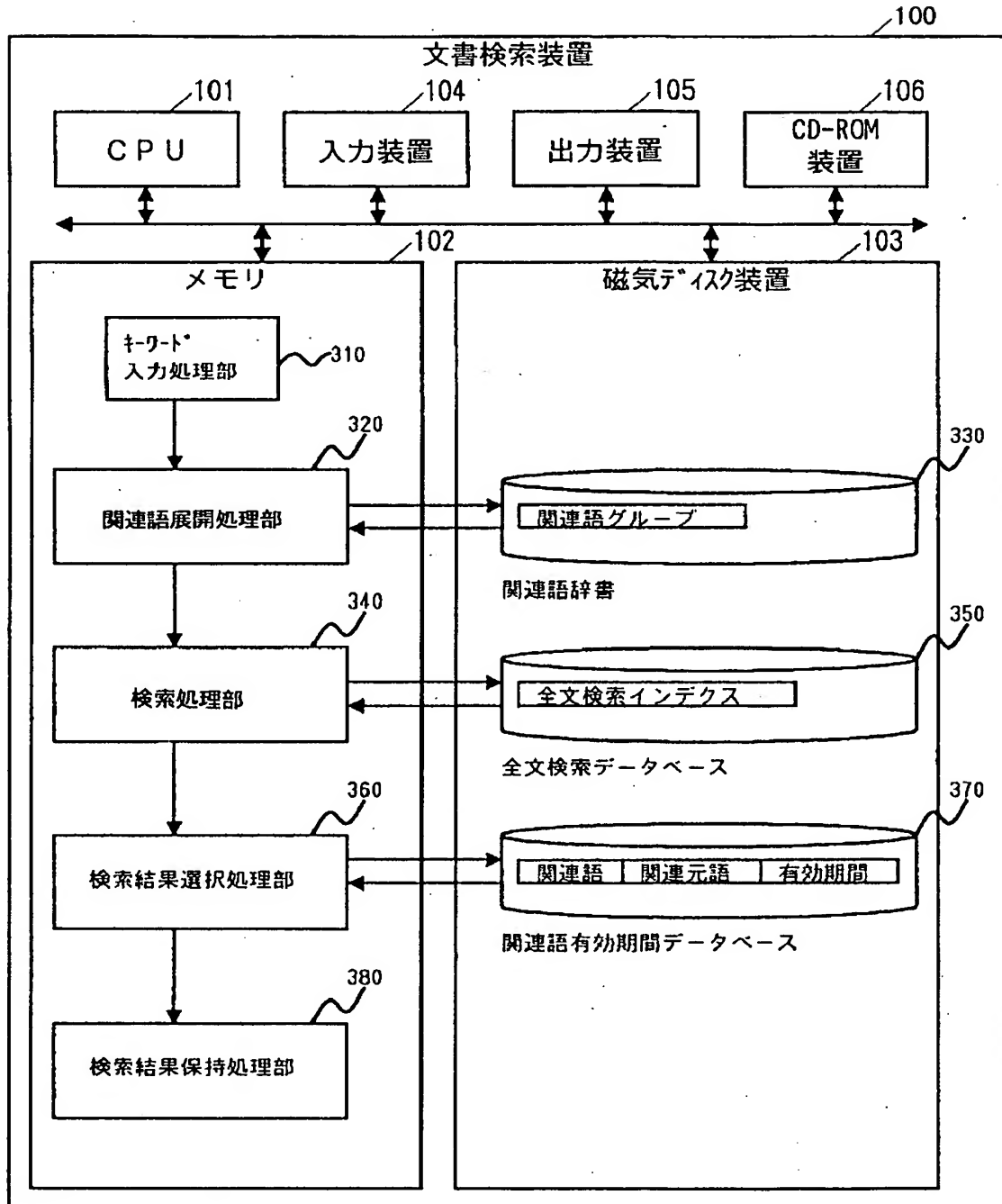
【図 6】

図 6



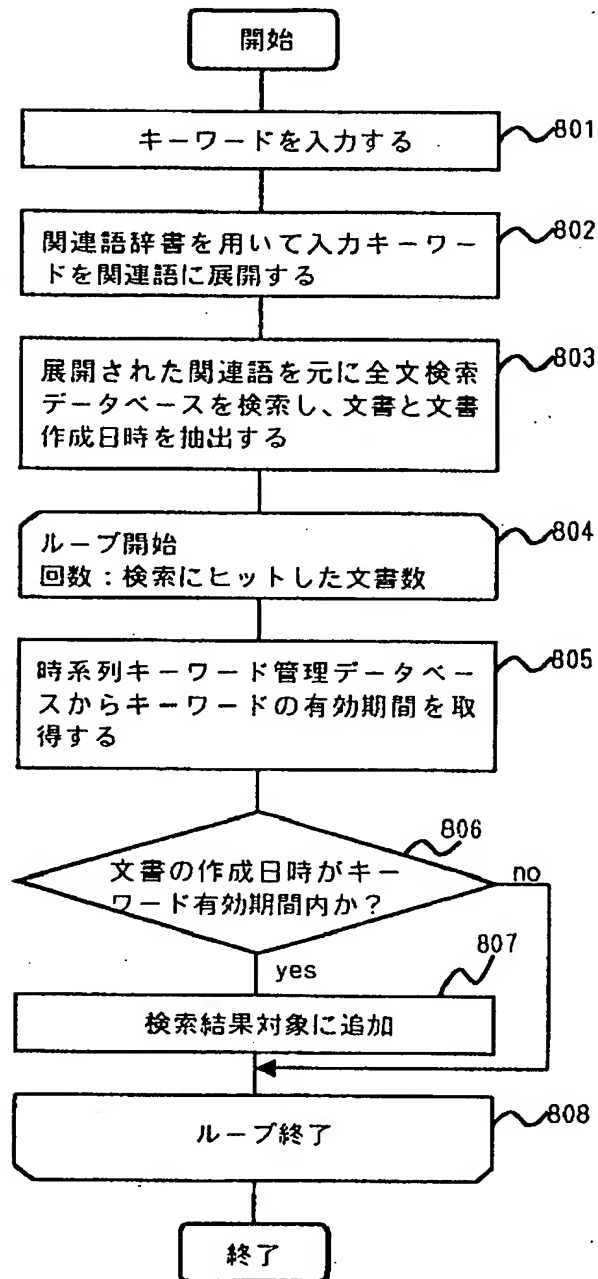
【図 7】

図 7

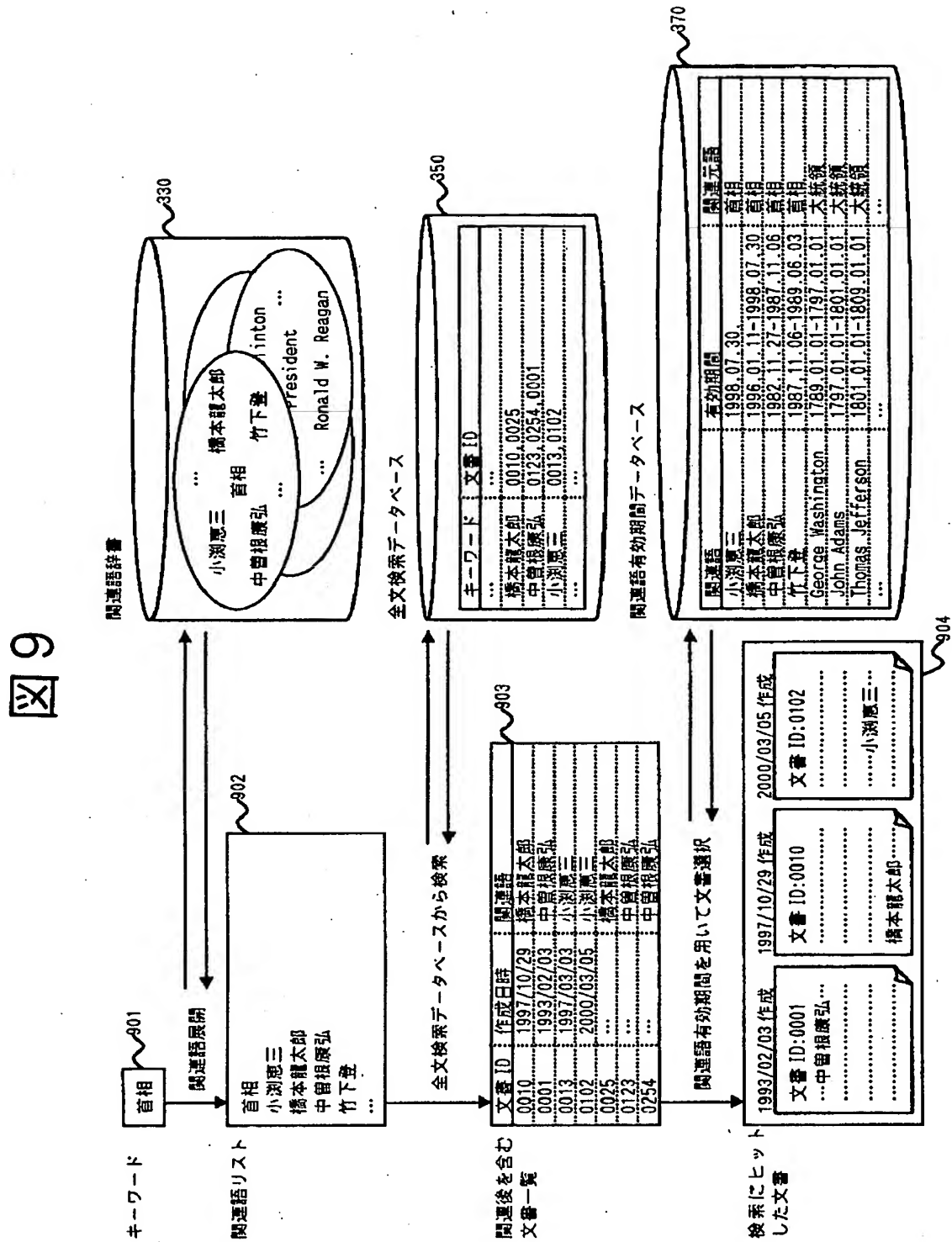


【図 8】

図 8



【图9】



【書類名】 要約書

【要約】

【課題】 ユーザの意図に合った適切な関連語を検索して文書検索作業の効率を向上させることが可能な技術を提供する。

【解決手段】 キーワードを用いて文書データベースから所望の文書を検索する文書検索方法において、入力されたキーワードに関連する関連語とその関連語の有効期間を抽出するステップと、前記抽出した関連語を検索語として文書の検索を行うステップと、前記抽出した有効期間内の文書を前記検索された文書の中から選択するステップとを有するものである。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願2001-002810
受付番号	50100019399
書類名	特許願
担当官	第七担当上席 0096
作成日	平成13年 1月11日

<認定情報・付加情報>

【提出日】	平成13年 1月10日
-------	-------------

出 願 人 履 歴 情 報

識別番号

[000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所